BACK

Principle Component Analysis for Uniqueness in Puget Sound Hydrographic Stations (1989-2003) – Let the Data Speak!

S L Albertson (Washington State Dept. of Ecology, 300 Desmond Drive POB 47710, Olympia, WA 98504-7710; 360-407-6676; alberts@ocean.washington.edu)

Keywords: Principle Component Analysis, Puget Sound, hydrography, EOFs

Abstract

The Washington State Department of Ecology (Ecology) samples approximately 40 hydrographic stations each month for complete vertical profiles by seaplane as part of the Puget Sound Ambient Monitoring Program (PSAMP). We visit some of these stations every year (core) and others only one year in three (rotational). Given a limited amount of resources (funding) how should stations be prioritized in terms of their importance? EPA's EMAP program uses a random selection process in choosing station locations, although typically oceanographers try to identify and monitor water mass endpoints (e.g., large rivers, ocean signals); assuming that we have over-sampled, which stations are truly the most independent if we analyze the historic data itself?

I use monthly hydrographic (temperature, salinity), dissolved oxygen, chlorophyll and nutrient data from the Puget Sound over a 14-year period to investigate this question. In order to make my analysis, I produce a composite year and eliminate any station that we did not sample for at least three years total, in an attempt to remove interannual bias. I further composite these into seasons: winter (Nov., Dec., Jan.), spring (Feb., Mar., Apr.), summer (May, Jun., Jul.), and autumn (Aug., Sep., Oct.) and compare normalized (mean-centered and divided by their standard deviation) parameters using principle component analysis (see also empirical orthogonal functions) and cluster analysis (dendrograms, tree diagrams).

Introduction

The advent of Integrated Ocean Observing Systems (IOOS) under the auspices of Ocean.US gives impetus to revisit an analysis of the efficiency/redundancy in the present list of stations monitored by Ecology as part of the Puget Sound Ambient Monitoring Program (PSAMP). Ecology has visited about 50 stations (40/month for a given year) monthly by seaplane during the period 1989-2003 (Fig. 1); some routinely, some only for a few years. Although the data go back to 1973, Ecology began using more accurate methods (e.g., a CTD) in 1989 and I will restrict this analysis by only looking at data collected after that time, and only at stations for which there is at least a record of at least three years.

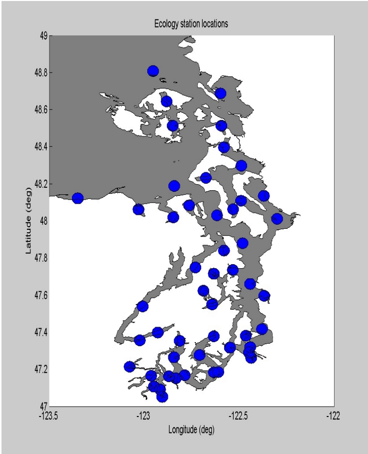


Figure 1. Stations visited from 1989-2003 with at least three years of data.

Methods

Temperature (T), salinity (S), nitrate (NO3-N, or N), chlorophyll *a* (C), and dissolved oxygen (O) are available from Ecology's website at http://www.ecy.wa.gov/programs/eap/mar_wat/mwm_intr.html. In order to use as many stations as possible, even those only sampled occasionally during this period, I had to composite the data. In so doing I give up a lot of important subtle vertical microstructure that might be important in determining processes like mixing (see Bos, Newton in this volume). I give up that variance in order to examine the mean seasonal trend, and to see which stations are most similar or unique for the mean annual trend using Principle Component Analysis (PCA). I mean-center and normalize (divide by the standard deviation) these data for statistical comparison.

I also composite these data into seasons according to a method first suggested by C. Ebbesmeyer (LOTT, 1998) during the Lacey Olympia Thurston Tumwater (LOTT) wastewater treatment plant recertification studies from 1996-7, by light level. Based on measured incident sunlight levels for 2004 obtained from the rooftop sensor at the atmospheric sciences department at UW (http://www-k12.atmos.washington.edu/k12/grayskies/nw_weather.html), I composite the data by season (Fig. 2).

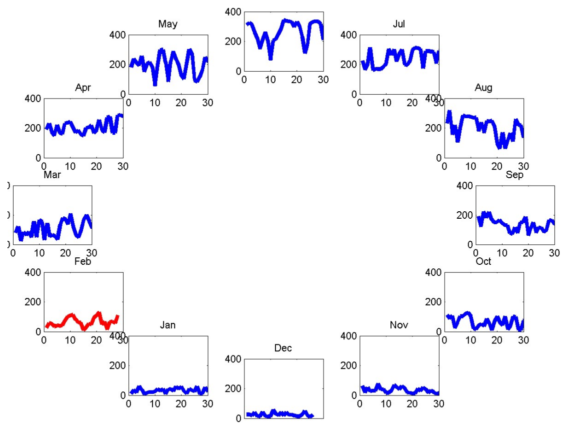


Figure 2. I use the incident solar energy, power (W/m^2) through time (day-of-month), to define seasons. Winter is N, D, J; spring is F, M, A and so on.

Results

Before discussing the statistical result of the principle component analysis (PCA) I interpret the oceanography of the mean seasonal cycle seen in the data. The physical distribution of water masses on a TS-diagram, from all the data prior to compositing, shows a fuzzy oceanic end-member, a salty summer end-member, and a winter freshet end-member (Fig. 3). Overlaying the composite dataset reveals several lobes that when compared to data from two recent Puget Sound Regional Synthesis Model (PRISM) cruises can be interpreted as South Puget Sound and Hood Canal summer and winter data, as well as those from the Whidbey Basin and Main Basin (Fig. 4). Although the PRISM data are from a warmer and drier year (2001), which is evident from the warmer and saltier location of the oceanic end-member on the TS-diagram, South Puget Sound and Hood Canal data outflank those from the north indicating a more extreme climate in both summer and winter.

BACK

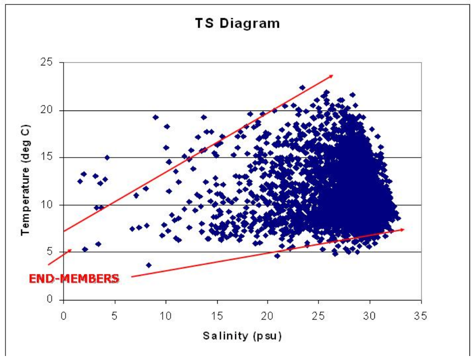


Figure 3. A TS-diagram of all data used prior to seasonal composite.

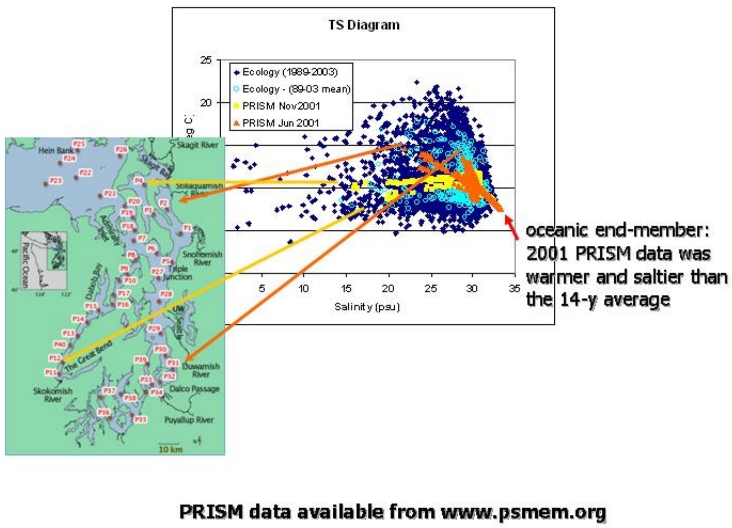


Figure 4. The TS-diagram, as before, overlying the composite-data as well as data from the summer (June) and winter (December) 2001 PRISM cruises.

An informed reader might expect that the highest summer nutrient levels at ambient (i.e., away from human sources) stations would occur at the oceanic end-member, where upwelling driven by offshore north wind drives nitrate-rich water landward via Ekman transport (Fig. 5).

BACK

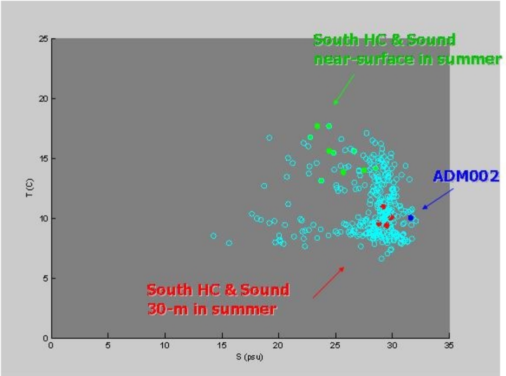


Figure 5. TS-diagram of composite data with oceanic end-member highlighted (blue) and stations with summer nutrient depletion at the surface (green) and the deep (30-m) data at those same stations (red).

However, many of the same stations that exhibit surface nutrient depletion in the summer have the highest levels of nitrate at depth, primarily those in the southern Hood Canal (Fig. 6). This is justification for including sinking of organic nitrogen and remineralization processes in Hood Canal models currently under development.

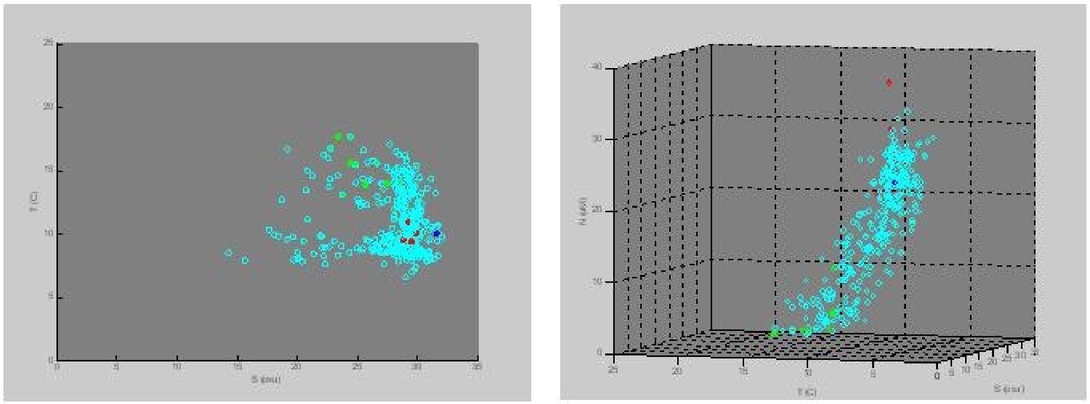


Figure 6. Stations in southern Hood Canal (red) have higher levels of nitrate at depth than those nearest to the ocean (blue).

Over-wintering phytoplankton are most prevalent in South Puget Sound and Hood Canal (Fig. 7a), with traceable amounts found in several other extremities such as the dead-end of East Sound in Orcas Island. Peak chlorophyll concentrations occur first in these areas then move north as the spring progresses into summer (Fig. 7b).

BACK

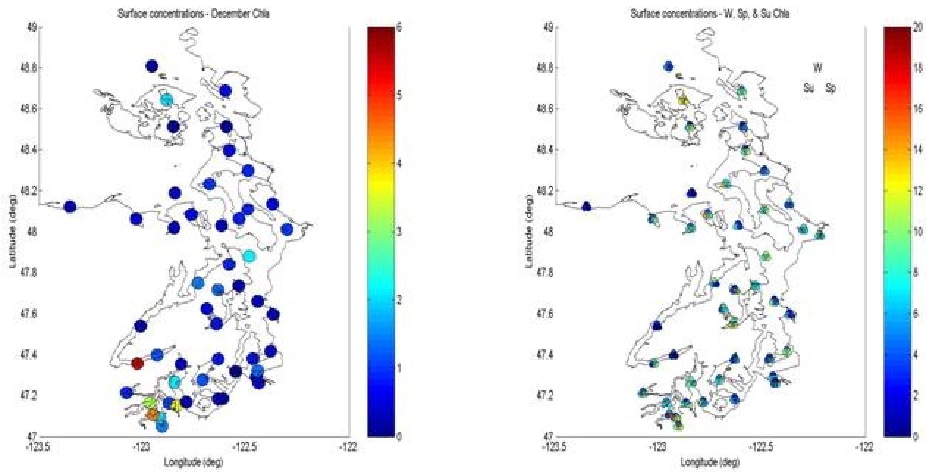
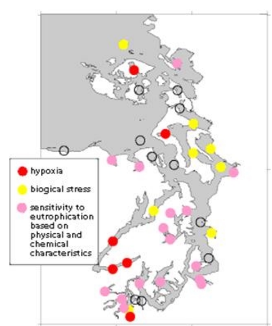


Figure 7. Seasonal chlorophyll a progression from a) over-winter and b) winter, spring, and summer.

DO levels are lowest in late summer to fall in Hood Canal, Budd Inlet, Penn Cove, and East Sound in Orcas Island although low DO occurs sporadically elsewhere in South Puget Sound and Whidbey Basin (Fig. 8).



http://www.ecy.wa.gov/programs/eap/mar_wat/eutrophication.html

Figure 8. Areas in Puget Sound with lowest DO concentrations.

A tree diagram of all the normalized data not surprisingly arranges stations next to each other. Stations with a lot of similarity in their seasonal cycle have shorter linkage distances between them (Fig. 9). Stations with the most similarity include Saratoga Passage (SAR003) and Possession Sound (PSS019), Nisqually (NSQ002) and Gordon Point (GOR001), and West Point (PSB003) and Elliott Bay (ELB015). Stations that are the most unique are interestingly enough near dead-ends (e.g., Totten Inlet (TOT001), Burley-Minter Lagoon (BML001), and East Sound (EAS001)).

BACK

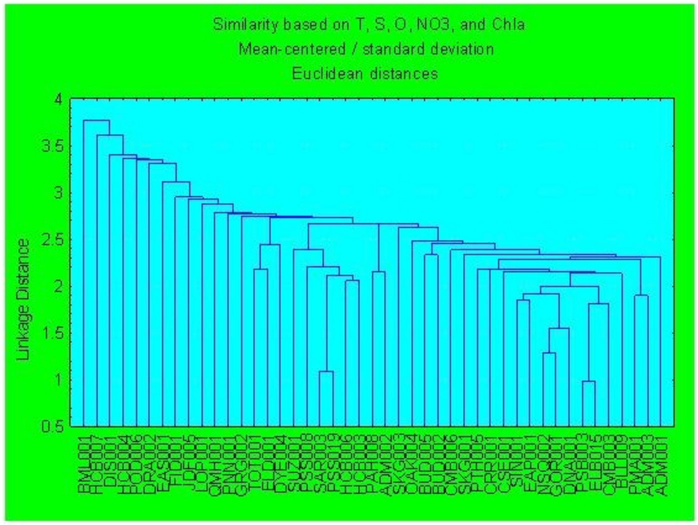


Figure 9. Tree diagram clustering most similar stations closest together.

The PCA again shows the importance of these dead-end stations. After a rotation to maximize the variation across the leading two factors (Fig. 10), the arrangement of stations into lobes corresponding to Basins is apparent. Although the Hood Canal and Whidbey Basin lobes overlap, little should be inferred from the similarity. Adding a third factor or some rotation in the present view provides separation. The explanation for this is that although the Skagit River (WB) and the Skokomish (HC) have similar hydrological cycles, the oxygen time series in Hood Canal is distinct and one could next expect to monitor both these basins with a single station.

The other noticeable feature of this PCA is that the end-member stations are toward the center of the diagram. This does not mean that they are not important to monitor, quite the opposite. They are not unique, but they are influential. All stations contain something from the end-members and that is why they are central in the PCA diagram.

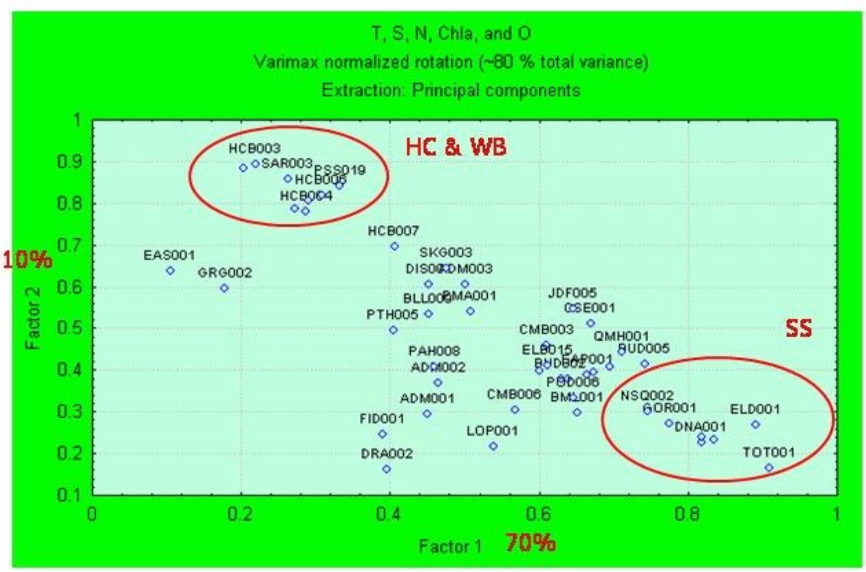


Figure 10. PCA of composite stations.

BACK

A caveat is that PCA has an underlying assumption that the data have oversampled the parameter space of interest; this is not necessarily the case. A single ambient station in an estuary at this scale is not necessarily located near the extreme conditions for any one of these variables (Fig. 11). Lowest summer oxygen may occur in one place, highest nitrate levels and salinity in another.

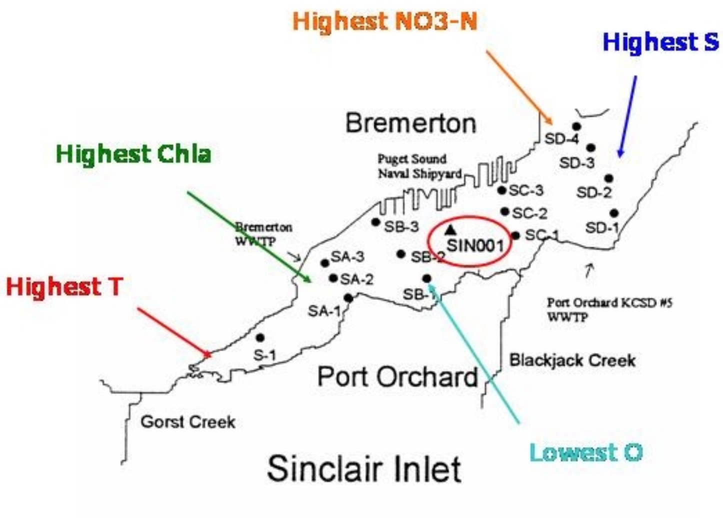


Figure 11. Separation of variable in a real estuary (Sinclair Inlet) relative to actual station.

Summary

Seasonal temperature and salinity ranges are more extreme to the south in South Puget Sound and Hood Canal. Stations with summer surface depletion of nitrate in Hood Canal also have the highest values of nitrate at depth indicating the importance of sinking and remineralization. Phytoplankton blooms, indicated by chlorophyll, occur landward (south) to seaward (north) in spring leaving a trail of depleted nutrients in their wake.

Stations should be prioritized by end-members (e.g., Strait of San Juan) and dead-end stations (e.g., HC, SS, Penn Cove, and East Sound (Orcas Island); most redundant are interior stations (e.g., GOR001). Station location for the Northwest Association of Networked Ocean Observing Systems (NANOOS, refer to www.nanoos.org) system will be determined by a series of factors such as obtaining boundary conditions for models, maintaining long time series, measuring key locations with water quality issues (e.g., 303d listings for DO), and model validation. Modeling and monitoring co-development is tantamount to understanding Puget Sound dynamics.

References

Albertson, S. Principle Component Analysis of Hydrographic Data in Sinclair and Budd Inlets; or How to Sample the Greatest Amounts of Variability (With the Least Effort). Puget Sound Research Proceedings, 1995.

Bos, this volume

LOTT Wastewater Management Partnership, Budd Inlet Scientific Study Final Report, prepared by Aura Nova Consultants, Brown and Caldwell, Inc., Evans-Hamilton, Inc., J.E. Edinger and Associates, Washington State Department of Ecology, and Dr. Alan Devol, University of Washington Department of Oceanography, August 1998.

Newton, this volume